

The Henchman Problem: Measuring Secrecy by the Minimum Distortion in a List

Curt Schieler and Paul Cuff

Abstract

We introduce a new measure of information-theoretic secrecy based on rate-distortion theory and study it in the context of the Shannon cipher system. Whereas rate-distortion theory is traditionally concerned with a single reconstruction sequence, in this work we suppose that an eavesdropper produces a list of 2^{nR_L} reconstruction sequences and measure secrecy by the minimum distortion over the entire list. We show that this setting is equivalent to one in which an eavesdropper must reconstruct a single sequence, but also receives side information about the source sequence and public message from a rate-limited henchman (a helper for an adversary). We characterize the optimal tradeoff of secret key rate, list rate, and eavesdropper distortion. The solution hinges on a problem of independent interest: lossy compression of a codeword drawn uniformly from a random codebook. We also characterize the solution to the lossy communication version of the problem in which distortion is allowed at the legitimate receiver. The analysis in both settings is greatly aided by a recent technique for proving source coding results with the use of a likelihood encoder.

I. INTRODUCTION

A ubiquitous model in information-theoretic secrecy is the Shannon cipher system [2] in which two nodes who share secret key want to communicate losslessly in the presence of an eavesdropper. As depicted in Figure 1, Node A views an i.i.d. source sequence X^n and uses the shared secret key K that is independent of the source to produce an encrypted message M . Node B uses the message and the key to produce \hat{X}^n . An eavesdropper views the message and knows the scheme that Nodes A and B employ.

Also ubiquitous is the investigation of how to measure secrecy when there is not enough key to ensure perfect secrecy, i.e. when the key rate is less than the entropy of the information source. One potential solution, proposed by Yamamoto in [3], is to measure secrecy by the distortion that an eavesdropper incurs in attempting to reconstruct

C. Schieler (schieler@princeton.edu) and P. Cuff (cuff@princeton.edu) are with the Department of Electrical Engineering, Princeton University, Princeton, NJ, 08544.

This work was supported in part by the National Science Foundation under Grants CCF-1116013 and by the Air Force Office of Scientific Research under Grant FA9550-12-1-0196. Part of this work was presented in [1].

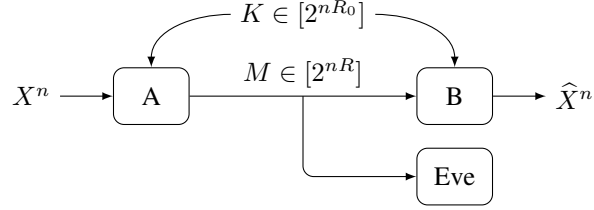


Fig. 1: The Shannon cipher system with secret key rate R_0 and communication rate R . In this paper, we measure secrecy by the minimum distortion in a list of reconstruction sequences $\{Z^n(1), \dots, Z^n(2^{nR_L})\}$ that the eavesdropper produces.

the source sequence. In accordance with the usual constructs in rate-distortion theory, this means that Nodes A and B want to maximize the following expression over all possible codes:

$$\min_{z^n(m)} \mathbb{P}[d(X^n, z^n(M)) \geq D]. \quad (1)$$

Although this seems like a reasonable objective at first glance, it was shown in [4] that simple codes employing negligible rates of secret key can force this probability to one, regardless of the distortion level D . The reason for this disconcerting result is that the accompanying secrecy guarantees can be fragile, as the following example elucidates. Let X^n be i.i.d. $\text{Bern}(1/2)$ and suppose that there is just one bit of secret key, i.e. $K \in \{0, 1\}$. Encrypt by transmitting X^n itself if $K = 0$ and X^n with all its bits flipped if $K = 1$. In this scenario, any optimal reconstruction Z^n that the eavesdropper produces has expected hamming distortion equal to $1/2$, the highest expected distortion that the eavesdropper could possibly incur. Despite this, the eavesdropper actually knows quite a bit about X^n , namely that it is one of two sequences. Indeed, the guarantee of secrecy is rather fragile because if the eavesdropper learns just one bit of the source sequence, then the entire sequence is compromised.

In view of the previous example, one way to strengthen a distortion-based measure of secrecy is to design schemes around the assumption that the eavesdropper has access to some side information. In [4], this is accomplished by supposing that eavesdropper views the causal behavior of the system; in particular, the eavesdropper reconstructs Z_i based on X^{i-1} and the public message M .

In this paper, we study another distortion-based approach to measuring secrecy in the Shannon cipher system. Instead of requiring a single reconstruction sequence Z^n , we suppose that the eavesdropper produces a list of 2^{nR_L} reconstructions $\{Z^n(1), \dots, Z^n(2^{nR_L})\}$ and consider the minimum distortion over the entire list. This is somewhat reminiscent of equivocation (i.e., the conditional entropy $H(X^n|M)$), which also purports to measure the uncertainty of the eavesdropper. However, an important difference in the measure we study is that the structure of the uncertainty is built directly into the definition. The eavesdropper's equivocation merely provides a lower bound on the size of the smallest list that contains the exact source sequence X^n . On the other hand, the optimal tradeoff between secret key rate, distortion, and list rate will give us a function $R_L(R_0, D)$ that precisely quantifies the size of the smallest list that an eavesdropper is able to produce that reliably contains a sequence of distortion

D.

Quantifying secrecy in terms of lists and distortion has been done previously in [5] and [6], where the eavesdropper is modeled as a “guessing wiretapper” who produces a sequence of reconstructions. After each estimate, the eavesdropper receives feedback about whether or not the reconstruction was within a certain distortion level.¹ As soon as the distortion level is reached, the eavesdropper stops guessing; the moments of the number of guesses needed indicate the secrecy of the system. Our approach differs from these works in that there is no sequential guessing (no testing mechanism) and the list size is fixed.

Organization

This paper considers the list-reconstruction measure of secrecy and establishes the information-theoretic characterization of the optimal tradeoffs among the secret key rate, list rate, and distortion at the eavesdropper. We divide the paper into two parts. First, we introduce and solve the problem when lossless communication is required between the legitimate parties (Sections II–V). We then introduce the lossy communication setting and solve the corresponding problem (Sections VI–VIII), reusing components from the preceding sections where possible. Although the lossy communication setting is a generalization of the lossless setting, there are several complications and subtleties that emerge that warrant the separation. For example, the converse proof is much more involved in the lossy setting.

In Section II, we formally define the list-based measure of secrecy and the lossless communication setting in which it will be first be analyzed. We also give an equivalent reformulation of the setting in terms of a malicious helper for the eavesdropper; the resulting “henchman problem” becomes the default formulation for the remainder of the paper. Section III contains Theorem 1, the characterization of the optimal tradeoffs in the lossless communication setting. The proof of Theorem 1 is presented in Section IV (converse) and Section V (achievability). In Section VI, we introduce the lossy communication version of the problem and characterize the optimal tradeoffs in Theorem 3. The converse and achievability proofs of Theorem 3 are given in Sections VII and VIII, respectively.

In addition to being a treatment of a new measure of secrecy for the Shannon cipher system, this paper is an endorsement of the efficacy of a likelihood encoder for proving source coding results. As detailed in [7], a likelihood encoder is a particular stochastic encoder which, when combined with a random codebook, manages to avoid many of the tedious and technical components of achievability proofs in lossy compression problems. The primary conduit for the analysis of a likelihood encoder is the “soft covering lemma”, which is expounded upon in [8]. In our case, the technique allows us to extract an idealized subproblem from the crucial part of the achievability proof and consider it independently of the original problem. The subproblem concerns the lossy compression of a codeword drawn uniformly from a random codebook.

¹In [5], the feedback concerns exact reconstructions, whereas [6] allows a distortion parameter.

II. PRELIMINARIES

A. Notation

All alphabets (e.g., \mathcal{X} , \mathcal{Y} , and \mathcal{Z}) are finite. The set $\{1, \dots, m\}$ is sometimes denoted by $[m]$. Given a per-letter distortion measure $d(x, z)$, we abuse notation slightly by defining

$$d(x^n, z^n) \triangleq \frac{1}{n} \sum_{i=1}^n d(x_i, z_i). \quad (2)$$

We also assume that for every $x \in \mathcal{X}$, there exists $z \in \mathcal{Z}$ such that $d(x, z) = 0$.

We denote the empirical distribution (or type) of a sequence x^n by T_{x^n} :

$$T_{x^n}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{x = x_i\}. \quad (3)$$

B. Total variation distance

Throughout the paper, we make frequent use of the total variation distance between two probability measures P and Q with common alphabet, defined by

$$\|P - Q\|_{\text{TV}} \triangleq \sup_{A \in \mathcal{F}} |P(A) - Q(A)|. \quad (4)$$

The following properties of total variation distance are quite useful.

Property 1. *Total variation distance satisfies:*

(a) *If the support of P and Q is a countable set \mathcal{X} , then*

$$\|P - Q\|_{\text{TV}} = \frac{1}{2} \sum_{x \in \mathcal{X}} |P(\{x\}) - Q(\{x\})|. \quad (5)$$

(b) *Let $\varepsilon > 0$ and let $f(x)$ be a function with bounded range of width $b > 0$. Then*

$$\|P - Q\|_{\text{TV}} < \varepsilon \implies |\mathbb{E}_P f(X) - \mathbb{E}_Q f(X)| < \varepsilon b, \quad (6)$$

where \mathbb{E}_P indicates that the expectation is taken with respect to the distribution P .

(c) *Let $P_X P_{Y|X}$ and $Q_X P_{Y|X}$ be two joint distributions with common channel $P_{Y|X}$. Then*

$$\|P_X P_{Y|X} - Q_X P_{Y|X}\|_{\text{TV}} = \|P_X - Q_X\|_{\text{TV}}. \quad (7)$$

(d) *Let P_X and Q_X be marginal distributions of P_{XY} and Q_{XY} . Then*

$$\|P_X - Q_X\|_{\text{TV}} \leq \|P_{XY} - Q_{XY}\|_{\text{TV}}. \quad (8)$$

C. Problem setup

As shown in Figure 1, Node A observes a source sequence X^n that is i.i.d. according to a distribution P_X . Nodes A and B share common randomness $K \in [2^{nR_0}]$ that is uniformly distributed and independent of X^n . Node A sends a message M to Node B over a noiseless channel at rate R .

Definition 1. An (n, R, R_0) code consists of:

$$\text{Encoder: } f : \mathcal{X}^n \times [2^{nR_0}] \rightarrow [2^{nR}] \quad (9)$$

$$\text{Decoder: } g : [2^{nR}] \times [2^{nR_0}] \rightarrow \mathcal{X}^n \quad (10)$$

The encoder and decoder can be stochastic (in which case they are denoted by $P_{M|X^n, K}$ and $P_{\hat{X}^n|M, K}$).

The encrypted communication (the message M) is overheard perfectly by an eavesdropper who produces a list $\mathcal{L}(M) \subset \mathcal{Z}^n$ and incurs the minimum distortion over the entire list:

$$\min_{z^n \in \mathcal{L}(M)} d(X^n, z^n). \quad (11)$$

Using the secret key and the noiseless channel, Nodes A and B want to communicate losslessly while ensuring that the eavesdropper's optimal strategy suffers distortion above a given level with high probability. The generalization to lossy communication begins in Section VI.

Definition 2. The tuple (R, R_0, R_L, D) is achievable if there exists a sequence of (n, R, R_0) codes such that the error probability $\mathbb{P}[X^n \neq \hat{X}^n]$ vanishes and, $\forall \varepsilon > 0$,

$$\min_{\mathcal{L}(m): |\mathcal{L}| \leq 2^{nR_L}} \mathbb{P} \left[\min_{z^n \in \mathcal{L}(M)} d(X^n, z^n) \geq D - \varepsilon \right] \xrightarrow{n \rightarrow \infty} 1. \quad (12)$$

Thus, we allow the eavesdropper to use any list-valued function $\mathcal{L} : \mathcal{M} \rightarrow \{\mathcal{Z}^n\}_1^{2^{nR_L}}$, provided the cardinality of the range satisfies $|\mathcal{L}| \leq 2^{nR_L}$. Furthermore, we assume that the eavesdropper knows the (n, R, R_0) code and the distribution P_X .

D. The henchman problem

So far, the problem has been formulated in terms of an eavesdropper who produces a list of 2^{nR_L} reconstructions. It turns out that we can relate this formulation to one in which an eavesdropper reconstructs a single sequence; this is accomplished by supplying the eavesdropper with a rate-limited helper (a henchman). As depicted in Figure 2, the eavesdropper receives nR_L bits of side information from a henchman who has access to the source sequence X^n and the public message M . Since the eavesdropper and henchman cooperate, this means that the eavesdropper effectively receives the best possible nR_L bits of side information about the pair (X^n, M) to assist in producing a single reconstruction sequence Z^n .

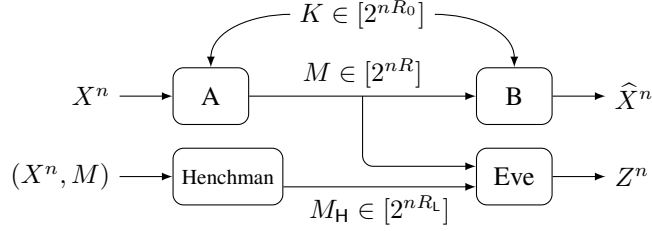


Fig. 2: The henchman problem. A rate-limited henchman has access to the source sequence and the public message. The eavesdropper produces a single reconstruction sequence Z^n based on the public message and the side information from the henchman.

Definition 3. The tuple (R, R_0, R_L, D) is achievable in the henchman problem if there exists a sequence of (n, R, R_0) codes such that the error probability $\mathbb{P}[X^n \neq \hat{X}^n]$ vanishes and, $\forall \varepsilon > 0$,

$$\min_{\substack{m_H(x^n, m), z^n(m, m_H): \\ |\mathcal{M}_H| \leq 2^{nR_L}}} \mathbb{P} \left[d(X^n, z^n(M, M_H)) \geq D - \varepsilon \right] \xrightarrow{n \rightarrow \infty} 1. \quad (13)$$

Thus, we allow the eavesdropper and henchman to jointly design a code consisting of an encoder $m_H(x^n, m)$ and a decoder $z^n(m, m_H)$, subject to the constraint $|\mathcal{M}_H| \leq 2^{nR_L}$. It can be shown that allowing a stochastic encoder or decoder does not decrease the eavesdropper's distortion. As in Definition 2, we assume that the adversarial entities are aware of the scheme that Nodes A and B employ, although this is not explicitly indicated in (13).

We now demonstrate the equivalence of the list reconstruction problem and the henchman problem.

Proposition 1. The tuple (R, R_0, R_L, D) is achievable in the list reconstruction problem if and only if it is achievable in the henchman problem. In other words, Definitions 2 and 3 are equivalent.

Proof: It is enough to show that the eavesdropper's scheme in the list reconstruction problem can be transformed to a scheme in the henchman problem that achieves the same (or less) distortion, and vice versa.

Let $\mathcal{L}(m)$ be the function that the eavesdropper uses to produce a list of reconstruction sequences. If the public message is M , the list $\mathcal{L}(M)$ can act as a codebook in the henchman problem. Knowing (X^n, M) , the henchman can transmit the index of the sequence in $\mathcal{L}(M)$ with the lowest distortion. Upon receiving the index and M , the eavesdropper reconstructs the corresponding sequence.

Conversely, suppose that the henchman and eavesdropper have devised an encoder $m_H(x^n, m)$ and a decoder $z^n(m, m_H)$. Upon observing the public message, the eavesdropper has a list of codewords (one for each m_H) that can be used for the list reconstruction problem. More precisely, the eavesdropper forms the list

$$\mathcal{L}(M) = \{z^n(M, m_H)\}_{m_H \in [2^{nR_L}]} \quad (14)$$

In both cases, it is straightforward to verify that the transformation maintains (or decreases) the distortion. To carry out the verification formally, it is enough to show that for any (n, R, R_0) code,

$$\min_{\mathcal{L}(m): |\mathcal{L}| \leq 2^{nR_L}} \mathbb{P} \left[\min_{z^n \in \mathcal{L}(M)} d(X^n, z^n) \geq D \right] = \min_{\substack{m_H(x^n, m), z^n(m, m_H): \\ |\mathcal{M}_H| \leq 2^{nR_L}}} \mathbb{P} \left[d(X^n, z^n(M, M_H)) \geq D \right]. \quad (15)$$

To show (\geq) , fix a list reconstruction function $\mathcal{L}(m)$ and define a henchman encoder and eavesdropper decoder by

$$m_H(x^n, m) = \arg \min_{j \in [2^{nR_L}]} d(x^n, \mathcal{L}(m, j)) \quad (16)$$

$$z^n(m, m_H) = \mathcal{L}(m, m_H), \quad (17)$$

where $\mathcal{L}(m, j)$ denotes the j th element of the list $\mathcal{L}(m)$. Then we have

$$\mathbb{P} \left[\min_{z^n \in \mathcal{L}(M)} d(X^n, z^n) \geq D \right] = \mathbb{P} \left[d(X^n, z^n(M, M_H)) \geq D \right] \quad (18)$$

$$\geq \min_{\substack{m_H(x^n, m), z^n(m, m_H): \\ |\mathcal{M}_H| \leq 2^{nR_L}}} \mathbb{P} \left[d(X^n, z^n(M, M_H)) \geq D \right]. \quad (19)$$

To show (\leq) , fix a henchman encoder $m_H(x^n, m)$ and eavesdropper decoder $z^n(m, m_H)$ and define a list reconstruction function by

$$\mathcal{L}(m) = \{z^n(m, m_H)\}_{m_H \in [2^{nR_L}]}. \quad (20)$$

Then we have

$$\mathbb{P} \left[d(X^n, z^n(M, M_H)) \geq D \right] \geq \mathbb{P} \left[\min_{z^n \in \mathcal{L}(M)} d(X^n, z^n) \geq D \right] \quad (21)$$

$$\geq \min_{\mathcal{L}(m): |\mathcal{L}| \leq 2^{nR_L}} \mathbb{P} \left[\min_{z^n \in \mathcal{L}(M)} d(X^n, z^n) \geq D \right]. \quad (22)$$

■

III. MAIN RESULT (LOSSLESS COMMUNICATION)

When lossless communication is required between the legitimate parties, we have the following characterization of the tradeoff among the communication rate, secret key rate, list rate (or henchman rate), and eavesdropper's distortion.

Theorem 1. *Given a source distribution P_X and a distortion function $d(x, z)$, the closure of achievable tuples (R, R_0, R_L, D) is the set of tuples satisfying*

$$R \geq H(X) \quad (23)$$

$$D \leq D(R_L) \cdot \mathbf{1}\{R_0 > R_L\},$$

where $D(\cdot)$ is the point-to-point distortion-rate function:

$$D(R) \triangleq \min_{P_{Z|X}: R \geq I(X; Z)} \mathbb{E}[d(X, Z)]. \quad (24)$$

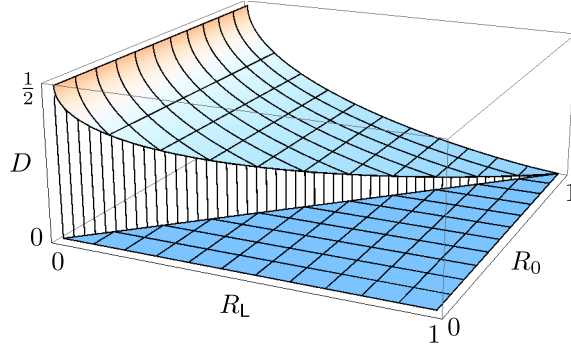


Fig. 3: The region in Theorem 1 for source distribution $P_X \sim \text{Bern}(1/2)$ and distortion measure $d(x, z) = \mathbf{1}\{x \neq z\}$.

Perhaps the most striking part of Theorem 1 is that the region is discontinuous. Fixing a rate of secret key R_0 , observe that when the list rate R_L is strictly less than R_0 , the (R_L, D) tradeoff follows the point-to-point rate-distortion function. However, as soon as R_L equals or exceeds the secret key rate, the eavesdropper's distortion drops to zero (the minimum distortion possible) because all possible decryptions can be enumerated in a list of size 2^{nR_0} . Figure 3 illustrates Theorem 1 for a $\text{Bern}(1/2)$ source and hamming distortion; the communication rate is assumed to satisfy $R \geq H(X)$ and has no effect on the (R_0, R_L, D) tradeoff.

Note that setting $R_L = 0$ in the region of Theorem 1 corresponds to requiring a single reconstruction (without a henchman), which was Yamamoto's original formulation of the problem in [3]. In this case, we see that any positive rate of secret key results in distortion $D(0)$, the maximum expected distortion that can occur.

In the context of the list reconstruction formulation, Theorem 1 implies that when Nodes A and B act optimally and $R_L < R_0$, the eavesdropper's best strategy is to simply ignore the public message and list the codewords from a good point-to-point rate-distortion codebook. In particular, the public message is useless to the eavesdropper in this regime. However, when $R_L \geq R_0$, the eavesdropper uses a different strategy and produces all possible decryptions of the public message. When we consider the lossy communication setting, we will see that a similar strategy switch occurs.

We now prove the achievability and converse portions of Theorem 1. For the entirety of the proof, we use the henchman formulation instead of the list reconstruction one. The main idea in the proof of achievability concerns the problem of compressing codewords from a random codebook beyond the rate-distortion limit; the proof also relies on a likelihood encoder [7] and the soft covering lemma [8, Lemma IV.1]. The converse is straightforward, as we now show.

IV. CONVERSE (LOSSLESS COMMUNICATION)

The constraint $R \geq H(X)$ is a consequence of the lossless source coding theorem. The constraint on D splits into two cases depending on the relation between R_0 and R_L . If $R_L \geq R_0$, then any scheme that Nodes A and B use to achieve lossless compression can be exploited by the eavesdropper and the henchman. Since they can both enumerate the 2^{nR_0} possible decryptions of M , the henchman can simply send the index of the correct decryption, which results in zero distortion. On the other hand, if $R_L < R_0$ then the eavesdropper and the henchman can ignore M altogether and simply use a point-to-point rate-distortion code to describe X^n within distortion $D(R_L)$. Therefore, regardless of the code that Alice and Bob use for lossless communication, the eavesdropper and the henchman can achieve distortion less than or equal to $D(R_L) \cdot \mathbf{1}\{R_0 > R_L\}$.

V. ACHIEVABILITY (LOSSLESS COMMUNICATION)

Viewing the problem from the perspective of the adversarial entities, we see that the henchman observes the pair (X^n, M) and encodes a message M_H , and the eavesdropper observes (M, M_H) and decodes Z^n ; their goal is to minimize the distortion $d(X^n, Z^n)$. This describes the usual rate-distortion setting with additional information M available at encoder and decoder. In other words, for a given $M = m$, the henchman observes a sequence X^n drawn from a source distribution $P_{X^n|M=m}$ and describes the sequence to the eavesdropper using a rate-limited channel; the conditional distribution $P_{X^n|M=m}$ is the effective source distribution because both the henchman and the eavesdropper know the public message.

Observing that $P_{X^n|M=m}$ is induced entirely by the actions of Node A, let us assume for the moment that Node A uses the following random binning scheme to encode the source sequence. First, randomly divide the set of typical x^n sequences into bins of size 2^{nR_0} . This binning is known to everyone, including the adversaries. To encode X^n , Node A transmits the message $M = (M_p, M_s)$, where M_p is the bin containing X^n , and M_s is the index within that bin, one-time padded with K . Note that the one-time pad renders M_s statistically independent of X^n and M_p . Thus, for this choice of encoder, the induced distribution $P_{X^n|M}$ corresponds to choosing a sequence roughly uniformly at random from bin M_p (because of the asymptotic equipartition property). Furthermore, the asymptotic equipartition property and the randomness of the binning suggest that the 2^{nR_0} sequences in bin M_p were approximately chosen i.i.d. according to $\prod_{i=1}^n P_X(x_i)$. Therefore, very roughly speaking, the random binning scheme results in a distribution $P_{X^n|M=m}$ that corresponds to selecting a sequence uniformly from a random codebook whose codewords are generated independently and identically according to $\prod_{i=1}^n P_X(x_i)$. If this is true, then the joint goal of the henchman and the eavesdropper becomes the following: lossy compression (at rate R_L) of a codeword drawn uniformly from a random codebook of size 2^{nR_0} . We now delve into this subproblem, the conclusion of which is the following: if $R_L < R_0$, then with high probability it is impossible to achieve distortion less than $D(R_L)$.

A. Interlude: lossy compression of a codeword drawn uniformly from a random codebook

Consider a codebook $c_x = \{x^n(1), \dots, x^n(2^{nR_c})\}$ consisting of 2^{nR_c} sequences. Select a codeword uniformly at random from c_x and denote it by $x^n(J)$, where $J \sim \text{Unif}[2^{nR_c}]$. An encoder describes $x^n(J)$ using a noiseless link of rate R , and a decoder estimates it with a reconstruction sequence Z^n . Both the encoder and decoder know the codebook c_x . Notice that the relationship between this setting and the standard rate-distortion framework is that the input space is contracted from \mathcal{X}^n to a codebook c_x , and the source sequence is chosen uniformly at random from c_x instead of i.i.d. according to a distribution P_X .

Definition 4. For a fixed codebook $c_x \subseteq \mathcal{X}^n$, define an (n, c_x, R) code as an encoder $f : \mathcal{X}^n \times c_x \rightarrow [2^{nR}]$ and a decoder $g : [2^{nR}] \times c_x \rightarrow \mathcal{Z}^n$.

For a given D and fixed codebook c_x , the encoder and decoder want to maximize the probability that the distortion between $x^n(J)$ and Z^n is less than D :

$$\max_{(n, c_x, R) \text{ codes}} \mathbb{P}[d(x^n(J), Z^n) \leq D]. \quad (25)$$

Instead of considering this objective for arbitrary codebooks, we generate a random codebook \mathcal{C}_x in which the codewords are drawn independently, each according to $\prod_{i=1}^n P_X(x_i)$. The setup is depicted in Figure 4.

In some regimes, the expression in (25) approaches one as blocklength increases. For example, if $R \geq R_c$ then the encoder can simply send the index of $X^n(J)$ within the codebook \mathcal{C}_x , thus ensuring zero distortion. Another example is when $R \geq R(D)$, in which case distortion D is achievable even without knowledge of \mathcal{C}_x , because $X^n(J)$ is i.i.d. according to P_X .

The regime we are interested in is when $R < R_c$ and $R < R(D)$. In this case, we find that with high probability (over the random codebook) it is impossible to achieve distortion D , i.e., the expression in (25) vanishes.

Theorem 2. Fix R , R_c and D . Let \mathcal{C}_x be a random codebook of 2^{nR_c} codewords, each drawn independently according to $\prod_{i=1}^n P_X(x_i)$. Let τ_n be any sequence that converges to zero sub-exponentially fast (i.e., $\tau_n = 2^{-o(n)}$). If

$$R < \min\{R(D), R_c\}, \quad (26)$$

then

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\mathcal{C}_x} \left[\max_{(n, \mathcal{C}_x, R) \text{ codes}} \mathbb{P}[d(X^n(J), Z^n) \leq D] > \tau_n \right] = 0. \quad (27)$$

Proof:

We first provide a brief, informal sketch of the proof idea. For an optimal (n, \mathcal{C}_x, R) code, there are on average $2^{n(R_c - R)}$ codewords in \mathcal{C}_x that map to each of the 2^{nR} reconstruction sequences in \mathcal{Z}^n . However, for a given

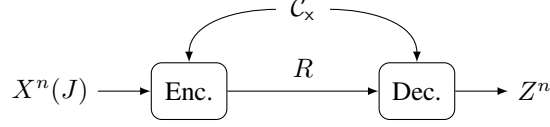


Fig. 4: Lossy compression of a codeword drawn uniformly from a random codebook $\mathcal{C}_x = \{X^n(1), \dots, X^n(2^{nR_c})\}$. Both the encoder and decoder know the codebook \mathcal{C}_x , and the encoder must describe a randomly chosen codeword $X^n(J)$, where $J \sim \text{Unif}[2^{nR_c}]$.

reconstruction sequence z^n , there are only (on average) $2^{n(R_c - R(D))}$ sequences in \mathcal{C}_x within distortion D of z^n , because the probability of an i.i.d sequence X^n being within distortion D of z^n is roughly $2^{-nR(D)}$. Since $2^{n(R_c - R(D))}$ is much smaller than $2^{n(R_c - R)}$, the probability that z^n yields distortion less than D is vanishingly small. In fact, this probability decays doubly exponentially, which means that the entire suite of 2^{nR} reconstruction sequences simultaneously yields distortion greater than D with high probability. In other words, the optimal code gives rise to distortion greater than D with high probability, which is what we want to show.

The first step is to restrict $X^n(J)$ to the δ -typical set $\mathcal{T}_\delta^n(X)$ by writing

$$\mathbb{P}[d(X^n(J), Z^n) \leq D] \leq \mathbb{P}[d(X^n(J), Z^n) \leq D, \mathcal{A}] + \mathbb{P}[\mathcal{A}^c], \quad (28)$$

where \mathcal{A} denotes the event $\{X^n(J) \in \mathcal{T}_\delta^n\}$. The δ -typical set is defined according to the notion of strong typicality:

$$\mathcal{T}_\delta^n(X) \triangleq \{x^n \in \mathcal{X}^n : \|T_{x^n} - P_X\|_{\text{TV}} < \delta\}, \quad (29)$$

where T_{x^n} denotes the empirical distribution (i.e., the type) of x^n . We will choose an appropriate δ later. Note that the second term in (28) vanishes in the limit for any $\delta > 0$ since $X^n(J)$ is i.i.d. according to P_X .

Although we defined a (n, \mathcal{C}_x, R) code as an encoder-decoder pair (f, g) , we will benefit from viewing a code as the combination of a codebook of z^n sequences and an encoder that is optimal for that codebook. In other words, treat an (n, \mathcal{C}_x, R) code as a codebook $c_z \subseteq \mathcal{Z}^n$ of size 2^{nR} , together with an encoder that maps $x^n \in \mathcal{C}_x$ to the $z^n \in c_z$ with the lowest distortion $d(x^n, z^n)$. This allows us to write

$$\begin{aligned} & \max_{(n, \mathcal{C}_x, R) \text{ codes}} \mathbb{P}[d(X^n(J), Z^n) \leq D, \mathcal{A}] \\ &= \max_{c_z(\mathcal{C}_x)} \mathbb{P}\left[\min_{z^n \in c_z(\mathcal{C}_x)} d(X^n(J), z^n) \leq D, \mathcal{A}\right], \end{aligned} \quad (30)$$

where the notation $c_z(\mathcal{C}_x)$ emphasizes that c_z is a function of the random codebook \mathcal{C}_x ; for simplicity, we suppress the n and R parameters of $c_z(\mathcal{C}_x)$.

Now we apply a union bound to the right-hand side of (30) and write

$$\mathbb{P}\left[\min_{z^n \in \mathcal{C}_x} d(X^n(J), z^n) \leq D, \mathcal{A}\right] \stackrel{(a)}{\leq} \sum_{z^n \in \mathcal{C}_x} \mathbb{P}\left[d(X^n(J), z^n) \leq D, \mathcal{A}\right] \quad (31)$$

$$\leq 2^{nR} \max_{z^n \in \mathcal{C}_x} \mathbb{P}\left[d(X^n(J), z^n) \leq D, \mathcal{A}\right] \quad (32)$$

$$\leq 2^{nR} \max_{z^n \in \mathcal{Z}^n} \mathbb{P}\left[d(X^n(J), z^n) \leq D, \mathcal{A}\right] \quad (33)$$

$$\stackrel{(b)}{=} 2^{-n(R_C-R)} \max_{z^n \in \mathcal{Z}^n} \sum_{j=1}^{2^{nR_C}} \mathbf{1}\{d(X^n(j), z^n) \leq D, X^n(j) \in \mathcal{T}_\delta^n\}, \quad (34)$$

where step (a) is a union bound, and step (b) uses the fact that $X^n(J)$ is chosen uniformly from \mathcal{C}_x . Notice that for a fixed z^n , the terms in the sum in (34) are i.i.d. random variables (due to the nature of the random codebook construction), which we henceforth denote by ξ_{j,z^n} :

$$\xi_{j,z^n} \triangleq \mathbf{1}\{d(X^n(j), z^n) \leq D, X^n(j) \in \mathcal{T}_\delta^n\}, \quad j = 1, \dots, 2^{nR_C}. \quad (35)$$

Using the equality in (30) and the bound in (34), we have

$$\begin{aligned} & \mathbb{P}\left[\max_{(n, \mathcal{C}_x, R) \text{ codes}} \mathbb{P}[d(X^n(J), Z^n) \leq D, \mathcal{A}] > \tau_n\right] \\ & \leq \mathbb{P}\left[\max_{z^n \in \mathcal{Z}^n} \sum_{j=1}^{2^{nR_C}} \xi_{j,z^n} > \tau_n 2^{n(R_C-R)}\right] \end{aligned} \quad (36)$$

$$\stackrel{(a)}{\leq} |\mathcal{Z}|^n \max_{z^n \in \mathcal{Z}^n} \mathbb{P}\left[\sum_{j=1}^{2^{nR_C}} \xi_{j,z^n} > \tau_n 2^{n(R_C-R)}\right], \quad (37)$$

where (a) is a union bound. If we can show that the probability in (37) decays doubly exponentially fast with n , then the proof will be complete. To that end, we first use a standard application of the method of types [9] to establish a bound on the expected value of ξ_{j,z^n} in the following lemma. The proof is relegated to the appendix.

Lemma 1. *If X^n is i.i.d. according to P_X , then for any z^n ,*

$$\mathbb{P}[d(X^n, z^n) \leq D, X^n \in \mathcal{T}_\delta^n] \leq 2^{-n(R(D)-o(1))}, \quad (38)$$

where $R(D)$ is the point-to-point rate-distortion function for P_X , and $o(1)$ is a term that vanishes as $\delta \rightarrow 0$ and $n \rightarrow \infty$.

From Lemma 1, we see that the expected value of $\sum_{j=1}^{2^{nR_C}} \xi_{j,z^n}$ is bounded above by approximately $2^{n(R_C-R(D))}$. Moreover, since a condition of the theorem being proved is that $R < R(D)$, it follows that $\tau_n 2^{n(R_C-R)}$ is exponentially larger than $2^{n(R_C-R(D))}$. Therefore, (37) is concerned with the probability that a sum of 2^{nR_C} i.i.d. Bernoulli random variables is exponentially far away from its mean. Such a probability decays at a doubly exponential rate, as the following Chernoff bound will imply.

Lemma 2. *If X^m is a sequence of i.i.d. $\text{Bern}(p)$ random variables, then*

$$\mathbb{P}\left[\sum_{i=1}^m X_i > k\right] \leq \left(\frac{e \cdot m \cdot p}{k}\right)^k. \quad (39)$$

Proof: The proof follows some of the usual steps for establishing Chernoff bounds.

$$\mathbb{P}\left[\sum_{i=1}^m X_i > k\right] \leq \min_{\lambda > 0} e^{-\lambda k} \prod_{i=1}^m \mathbb{E}[e^{\lambda X_i}] \quad (40)$$

$$= \min_{\lambda > 0} e^{-\lambda k} (p \cdot e^\lambda + 1 - p)^m \quad (41)$$

$$\leq \min_{\lambda > 0} e^{-\lambda k} (p \cdot e^\lambda + 1)^m \quad (42)$$

$$\leq \min_{\lambda > 0} e^{-\lambda k} e^{mpe^\lambda} \quad (43)$$

Substituting the minimizer $\lambda^* = \ln(\frac{k}{mp})$ gives the desired bound. \blacksquare

Using the bound on $\mathbb{E}[\xi_{j,z^n}]$ from Lemma 1, we can apply Lemma 2 to the probability in (37) by identifying

$$m = 2^{nR_C} \quad (44)$$

$$p \leq 2^{-n(R(D)-o(1))} \quad (45)$$

$$k = \tau_n 2^{n(R_C-R)}. \quad (46)$$

This gives

$$\mathbb{P}\left[\sum_{j=1}^{2^{nR_C}} \xi_{j,z^n} > \tau_n 2^{n(R_C-R)}\right] \leq 2^{-n\alpha 2^{n\beta}}, \quad (47)$$

where

$$\alpha = R(D) - R - o(1) \quad (48)$$

$$\beta = R_C - R - o(1). \quad (49)$$

For small enough δ and large enough n , both α and β are positive and bounded away from zero, and (47) vanishes doubly exponentially fast. Consequently, the expression in (37) vanishes, completing the proof of Theorem 2. \blacksquare

One can readily establish the following corollary to Theorem 2.

Corollary 1. *If $R < R_C$ and $R < R(D)$, then*

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}_x} \left[\min_{(n, C_x, R) \text{ codes}} \mathbb{P}[d(X^n(J), Y^n) \geq D] \right] = 1. \quad (50)$$

The interlude is now complete, and we can return to the achievability proof of Theorem 1.

B. Likelihood encoder

Earlier, we asserted that a scheme similar to random binning might give rise to an induced distribution $P_{X^n|M=m}$ that could be approximated by drawing a codeword uniformly from a random codebook. Then we could apply Corollary 1 to our problem by identifying (R_C, R) with (R_0, R_L) . Although it is possible that an encoder using random binning might yield this distribution, we turn instead to a likelihood encoder with a random codebook because it brings considerable clarity to the induced distributions involved.

Consider a codebook $c = \{x^n(m, k)\}$ consisting of $2^{n(R+R_0)}$ sequences from \mathcal{X}^n . The likelihood encoder of [7] for lossless reconstruction and for this codebook is a stochastic encoder defined by

$$P_{M|X^n K}(m|x^n, k) \propto \prod_{i=1}^n \mathbf{1}\{x_i = x_i(m, k)\}, \quad (51)$$

where \propto indicates that appropriate normalization is required.² The merit of using a likelihood encoder with a random codebook is that the resulting system-induced joint distribution of (X^n, M, K) , namely $P_{X^n M K} = P_{X^n} P_K P_{M|X^n K}$, can be shown to be close to an idealized distribution $Q_{X^n M K}$ defined by

$$Q_{X^n M K}(x^n, m, k) \triangleq 2^{-n(R+R_0)} \prod_{i=1}^n \mathbf{1}\{x_i = x_i(m, k)\}. \quad (52)$$

More precisely, one can use the soft covering lemma [8, Lemma IV.1] to prove the following.

Lemma 3. *Let $\mathcal{C} = \{X^n(m, k)\}, (m, k) \in [2^{nR}] \times [2^{nR_0}]$ be a random codebook with each codeword drawn independently according to $\prod_{i=1}^n P_X$. If $R > H(X)$, then*

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}} \|P_{X^n M K} - Q_{X^n M K}\|_{\text{TV}} = 0, \quad (53)$$

where the expectation is with respect to the random codebook and $\|\cdot\|_{\text{TV}}$ is total variation distance.

Proof: From the definition of $P_{X^n M K}$ and $Q_{X^n M K}$ we have $P_{M|X^n K} = Q_{M|X^n K}$. Using this fact, we have

$$\mathbb{E}_{\mathcal{C}} \|P_{X^n M K} - Q_{X^n M K}\|_{\text{TV}} \stackrel{(a)}{=} \mathbb{E}_{\mathcal{C}} \|P_{X^n K} - Q_{X^n K}\|_{\text{TV}} \quad (54)$$

$$= \mathbb{E}_{\mathcal{C}} \|P_{X^n} P_K - Q_{X^n|K} P_K\|_{\text{TV}} \quad (55)$$

$$= 2^{-nR_0} \sum_{k=1}^{2^{nR_0}} \mathbb{E}_{\mathcal{C}} \|P_{X^n} - Q_{X^n|K=k}\|_{\text{TV}}, \quad (56)$$

where (a) uses Property 1c. Since $R > H(X)$, the soft covering lemma implies that the summands vanish³:

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}} \|P_{X^n} - Q_{X^n|K=k}\|_{\text{TV}} = 0. \quad (57)$$

Without getting into the details of the soft covering lemma, it is worthwhile to briefly summarize the main idea. The lemma, which is expounded upon in [8], i The soft covering lemma applies to the current proof because $Q_{X^n|K=k}$

²In the rare case that no codeword is equal to the source sequence, an arbitrary index can be chosen.

³Furthermore, they vanish uniformly for all $k \in [2^{nR_0}]$.

is the output distribution induced by a memoryless channel acting on a random codebook of size 2^{nR} , and P_{X^n} is an i.i.d. distribution. Since we are considering lossless communication in this section, the relevant channel is the noiseless identity channel, and the relevant rate condition is $R > H(X)$. ■

Lemma 3 and the definition of total variation distance allow us to analyze the probability in (13) as if $Q_{X^n M K}$ were the true system-induced joint distribution instead of $P_{X^n M K}$. This is important because $Q_{X^n|M=m}$ is, as desired, uniform over a random codebook of size 2^{nR_0} :

$$Q_{X^n|M=m} = \text{Unif}\{X^n(m, 1), \dots, X^n(m, 2^{nR_0})\}. \quad (58)$$

To see the role of $Q_{X^n|M=m}$, first denote (for the sake of brevity) the event

$$\mathcal{E} = \{d(X^n, z^n(M, M_H)) \geq D(R_L) - \varepsilon\}. \quad (59)$$

Our objective is to show that when $R_L < R_0$, Nodes A and B can force the eavesdropper to incur distortion $D(R_L)$, i.e., there exists a sequence of codes that ensures (13).

Taking the expectation of (13) with respect to a random codebook, we have

$$\begin{aligned} & \mathbb{E}_{\mathcal{C}} \left[\min_{\substack{m_H(x^n, m), z^n(m, m_H): \\ |\mathcal{M}_H| \leq 2^{nR_L}}} \mathbb{P}_{P_{X^n M}}[\mathcal{E}] \right] \\ & \stackrel{(a)}{=} \mathbb{E}_{\mathcal{C}} \left[\min_{\substack{m_H(x^n, m), z^n(m, m_H): \\ |\mathcal{M}_H| \leq 2^{nR_L}}} \mathbb{P}_{Q_{X^n M}}[\mathcal{E}] \right] + o(1) \end{aligned} \quad (60)$$

$$= \mathbb{E}_{\mathcal{C}} \left[\min_{\substack{m_H(x^n, m), z^n(m, m_H): \\ |\mathcal{M}_H| \leq 2^{nR_L}}} \mathbb{E}[\mathbb{P}_{Q_{X^n M}}[\mathcal{E}|M]] \right] + o(1) \quad (61)$$

$$\stackrel{(b)}{=} \mathbb{E}_M \mathbb{E}_{\mathcal{C}} \left[\min_{\substack{m_H(x^n), z^n(m_H): \\ |\mathcal{M}_H| \leq 2^{nR_L}}} \mathbb{P}_{Q_{X^n|M}}[\mathcal{E}|M] \right] + o(1). \quad (62)$$

In step (a), we use Lemma 3 to change the underlying distribution from $P_{X^n M}$ to $Q_{X^n M}$ with vanishing penalty. Step (b) uses the fact that M and \mathcal{C} are independent under $Q_{X^n M}$. These steps bring us to the problem considered in the recent interlude: we must show that the henchman and the eavesdropper cannot design a code that achieves distortion $D(R_L)$ for the “source” $Q_{X^n|M=m}$.

Suppose that we are in the regime $R_L < R_0$. The expression in (62) is exactly what is addressed by Corollary 1, because the conditional distribution $Q_{X^n|M=m}$ corresponds to selecting a codeword uniformly from a random codebook of size 2^{nR_0} (as noted in (58)), and R_L is the rate of the message sent from the “encoder” (henchman) to the “decoder” (eavesdropper). Also, note that we are invoking the corollary with $D = D(R_L) - \varepsilon$; thus, $R_L < R(D)$ is satisfied. Hence, Corollary 1 gives

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}} \left[\min_{\substack{m_H(x^n), z^n(m_H): \\ |\mathcal{M}_H| \leq 2^{nR_L}}} \mathbb{P}_{Q_{X^n|M=m}}[\mathcal{E}|M=m] \right] = 1. \quad (63)$$

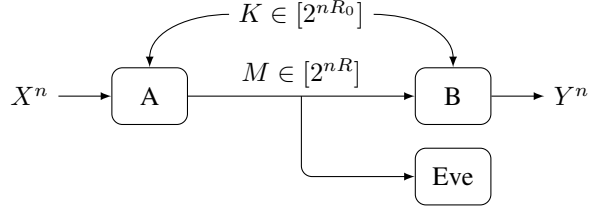


Fig. 5: Lossy communication. Secrecy is measured by the minimum distortion in a list of reconstruction sequences $\{Z^n(1), \dots, Z^n(2^{nR_L})\}$ that the eavesdropper produces. There are two distortion functions at play, $d_B(x, y)$ and $d_E(x, z)$.

The likelihood encoder also provides the required lossless communication between Nodes A and B; it is straightforward to show that (53) implies vanishing probability of error if the decoder is defined by $g(m, k) = x^n(m, k)$. Indeed,

$$\mathbb{E}_{\mathcal{C}} \mathbb{P}[X^n \neq \hat{X}^n] = \mathbb{E}_{\mathcal{C}} \mathbb{P}_{P_{X^n M K}}[X^n \neq X^n(M, K)] \quad (64)$$

$$\stackrel{(a)}{=} \mathbb{E}_{\mathcal{C}} \mathbb{P}_{Q_{X^n M K}}[X^n \neq X^n(M, K)] + o(1) \quad (65)$$

$$= 0 + o(1), \quad (66)$$

where step (a) follows from Lemma 3 and the definition of total variation distance.

We can conclude that there exists a codebook such that the associated likelihood encoder ensures (13) and lossless communication, because both hold when averaged over random codebooks. This completes the achievability portion of the proof of Theorem 1.

VI. LOSSY COMMUNICATION

We now generalize the problem to allow distortion at the legitimate receiver. As depicted in Figure 5, the receiver produces a reconstruction sequence Y^n , whose distortion is measured by $d(X^n, Y^n)$. Since there are two distortion measures (one for the receiver and one for the eavesdropper), we will distinguish them by using subscripts B and E.

Definition 5. The tuple (R, R_0, R_L, D_B, D_E) is achievable if there exists a sequence of (n, R, R_0) codes such that $\forall \varepsilon > 0$,

1) *Lossy communication:*

$$\mathbb{P}[d_B(X^n, Y^n) \leq D_B + \varepsilon] \xrightarrow{n \rightarrow \infty} 1. \quad (67)$$

2) *List secrecy:*

$$\min_{\mathcal{L}(m): |\mathcal{L}| \leq 2^{nR_L}} \mathbb{P}\left[\min_{z^n \in \mathcal{L}(M)} d_E(X^n, z^n) \geq D_E - \varepsilon\right] \xrightarrow{n \rightarrow \infty} 1. \quad (68)$$

Note that the decoder of a (n, R, R_0) code for lossy communication is a (possibly stochastic) decoder $P_{Y^n|MK}$.

As with the lossless version of problem, we can reformulate Definition 5 in terms of a rate-limited henchman. The henchman formulation for the lossy communication setting is exactly described by Figure 2 (with \hat{X}^n replaced by Y^n).

The optimal tradeoff between the various rates and distortions is the following.

Theorem 3. *Given a source distribution P_X and distortion functions $d_B(x, y)$ and $d_E(x, z)$, the closure of achievable tuples (R, R_0, R_L, D_B, D_E) is the set of tuples satisfying*

$$\begin{aligned} R &\geq I(X; Y) \\ D_B &\geq \mathbb{E} d_B(X, Y) \\ D_E &\leq \begin{cases} D(R_L) & \text{if } R_L < R_0 \\ \min\{D(R_L), D(R_L - R_0, P_{XY})\} & \text{if } R_L \geq R_0 \end{cases} \end{aligned} \quad (69)$$

for some $P_{XY} = P_X P_{Y|X}$, where $D(\cdot, P_{XY})$ is the point-to-point distortion-rate function with side information channel $P_{Y|X}$ to the encoder and decoder:

$$D(R, P_{XY}) \triangleq \min_{P_{Z|XY}: R \geq I(X; Z|Y)} \mathbb{E} d_E(X, Z). \quad (70)$$

When $R_L < R_0$, the eavesdropper's distortion is at least $D(R_L)$, just as it was when we considered lossless communication. This should not be surprising in light of the previous section, since less information is being revealed to the eavesdropper (the communication rate between Nodes A and B is lower). As before, the henchman can simply use a point-to-point rate-distortion code to achieve $D(R_L)$.

The more interesting regime is when $R_L \geq R_0$, i.e., the list rate (equivalently, the henchman's rate) is greater or equal to the rate of secret key. In this case, Theorem 3 says that a communication scheme can be designed such that the eavesdropper's distortion cannot be less than

$$\min\{D(R_L), D_Y(R_L - R_0)\}. \quad (71)$$

To see why these are the relevant distortions, consider the following. As we just mentioned, the henchman and the eavesdropper can always ignore the message M and use a point-to-point code to achieve $D(R_L)$. Alternatively, when $R_L \geq R_0$, the henchman can first use part of the rate R_L to communicate the secret key to the eavesdropper. Then, roughly speaking, the henchman and eavesdropper effectively share side information Y^n (since they both know M and K perfectly and can mimic the decoder), and can use the remaining rate $R_L - R_0$ to achieve distortion $D(R_L - R_0, P_{XY})$. Thus, one implication of Theorem 3 is that the henchman benefits from sending information about the secret key only if he describes it entirely; there is no benefit to communicating just part of the key to the eavesdropper.

VII. CONVERSE (LOSSY COMMUNICATION)

We now present the converse proof for Theorem 3. In the regime $R_0 > R_L$, the converse is the same as when we required lossless communication. Nodes A and B (the legitimate parties) cannot force distortion greater than $D(R)$ with high probability because the henchman and the eavesdropper can always ignore the public message M and simply use a good rate-distortion code to achieve distortion $D(R)$ with high probability. Note that this converse is “strong” in the sense that the probability of eavesdropper distortion being greater than $D(R)$ is not just bounded away from unity, it is actually vanishing. To be explicit, observe that if $R_L < R_0$ and $D_E > D(R)$, then the expression in (68) vanishes for all $\varepsilon < D_E - D(R)$. This follows from the achievability portion of point-to-point rate-distortion theory.

When $R_L \geq R_0$, the henchman’s rate is high enough that he can communicate the secret key to the eavesdropper and still have leftover rate $R_L - R_0$. Since the henchman and the eavesdropper both know M and K , they can mimic the decoder of Node B and produce side information Y^n . Notice that we have made two assumptions: the henchman knows the secret key, and the receiver uses a deterministic decoder. However, if the henchman were not able to determine the secret key exactly, then multiple keys would correspond to the same source sequence, which means that the decoder would effectively be stochastic. Thus, we are making just one assumption: that the decoder is deterministic. This assumption is valid because a stochastic decoder cannot be used to increase the eavesdropper’s distortion. Indeed, if we consider the list formulation of the problem, we see that eavesdropper’s performance is completely determined by X^n and M alone; the output of Node B does not play a role.⁴

So far, we have that the henchman and the eavesdropper share side information Y^n equal to the receiver’s reconstruction. Ideally, we would like to claim that (X^n, Y^n) are jointly i.i.d. according to some distribution P_{XY} and use the achievability portion of rate-distortion theory with side information at the encoder and decoder. Unfortunately, we cannot even claim that with high probability (X^n, Y^n) are jointly typical according to some P_{XY} because that is only guaranteed when Nodes A and B are using a nearly optimal rate-distortion code (i.e., one that operates near the rate-distortion tradeoff boundary). Instead, we rely on a different property of (X^n, Y^n) that will be given shortly in Lemma 5.

We will describe the henchman and eavesdropper’s scheme in terms of the joint type of (X^n, Y^n) ; to do this, we require the following straightforward extension of the type-covering lemma [10, Lemma 9.1] that accounts for side information (proof omitted). Regarding notation, \mathcal{T}_X^n denotes the set of sequences whose types coincide with a given distribution P_X , and $\mathcal{T}(\mathcal{X}^n)$ denotes the set of all joint types on sequences in \mathcal{X}^n .

Lemma 4. *Let $\tau > 0$ and $r \geq 0$. Fix a joint type $P_{XY} \in \mathcal{T}(\mathcal{X}^n \times \mathcal{Y}^n)$, and let $y^n \in \mathcal{T}_Y^n$. For $n \geq n_0(\tau)$, there exists a codebook $\mathcal{C}(y^n, P_{XY}) \subseteq \mathcal{Z}^n$ such that*

⁴Note that this would not be the case if we were considering distortion functions of the form $d_E(x, y, z)$ instead of $d_E(x, z)$.

1)

$$\frac{1}{n} \log |\mathcal{C}(y^n, P_{XY})| \leq r. \quad (72)$$

2) For all x^n such that $(x^n, y^n) \in \mathcal{T}_{XY}^n$,

$$\min_{z^n \in \mathcal{C}(y^n, P_{XY})} d(x^n, z^n) \leq D(r, P_{XY}) + \tau \quad (73)$$

We also require the following lemma from [11].

Lemma 5 ([11, Theorem 7]). *Consider any sequence of rate-distortion codes with rate $\leq R$. Then*

$$\limsup_{n \rightarrow \infty} I(T_{X^n Y^n}) \leq R \quad \text{a.s.}, \quad (74)$$

where $T_{X^n Y^n}$ denotes the type of (X^n, Y^n) and $I(\cdot)$ is the mutual information.

Now we can begin the converse proof for the regime $R_L \geq R_0$. Consider an achievable tuple (R, R_0, R_L, D_B, D_E) . By the same argument that was used in the regime $R_L < R_0$, we must have $D_E \leq D(R_L)$ because the henchman and the eavesdropper can always ignore the public message and use a good rate-distortion code to describe X^n .

Let $\varepsilon \in (0, 1/2)$. Define a set \mathcal{A}_n of joint distributions on $\mathcal{X} \times \mathcal{Y}$ by

$$\mathcal{A}_n \triangleq \left\{ \begin{array}{l} Q_{XY} : I_Q(X; Y) \leq R + \varepsilon \\ \mathbb{E}_Q d_B(X, Y) \leq D_B + \varepsilon \\ \|Q_X - P_X\|_{TV} \leq \varepsilon \end{array} \right\}. \quad (75)$$

We first show that

$$\lim_{n \rightarrow \infty} \mathbb{P}[T_{X^n Y^n} \in \mathcal{A}_n] = 1, \quad (76)$$

where $T_{X^n Y^n}$ denotes the type of (X^n, Y^n) . This can be proved by combining the following three facts:

1) From Lemma 5, we have

$$\lim_{n \rightarrow \infty} \mathbb{P}[I(T_{X^n, Y^n}) \leq R + \varepsilon] = 1. \quad (77)$$

2) From the definition of achievability and the equality $d_B(x^n, y^n) = \mathbb{E}_{T_{x^n y^n}} d_B(x, y)$, we have

$$\lim_{n \rightarrow \infty} \mathbb{P}[\mathbb{E}_{T_{X^n Y^n}} d_B(x, y) \leq D + \varepsilon] = 1. \quad (78)$$

3) From the weak law of large numbers, we have

$$\lim_{n \rightarrow \infty} \mathbb{P}[\|T_{X^n} - P_X\|_{TV} \leq \varepsilon] = 1. \quad (79)$$

With (76) in hand, choose n large enough so that

1) The eavesdropper cannot reconstruct with low distortion (this is an assumption of achievability):

$$\max_{\substack{m_H(x^n, m), z^n(m, m_H): \\ |\mathcal{M}_H| \leq 2^{nR_L}}}: \mathbb{P}[d_E(X^n, z^n(M, M_H)) < D_E - \varepsilon] < \varepsilon. \quad (80)$$

2)

$$\mathbb{P}[T_{X^n Y^n} \in A_n] \geq \varepsilon \quad (81)$$

3) Lemma 4 is satisfied with $\tau = \varepsilon$ and $r = R_L - R_0$.

4) The number of bits needed to express the joint type is negligible:

$$\frac{1}{n} |\mathcal{X}| |\mathcal{Y}| \log(n+1) \leq \varepsilon. \quad (82)$$

To compress x^n using side information y^n , the henchman first describes the joint type of (x^n, y^n) , then transmits the index of x^n in the codebook $\mathcal{C}(y^n, T_{x^n, y^n})$ that is guaranteed by Lemma 4. The description of the joint type only uses additional rate ε because the size of $\mathcal{T}(\mathcal{X} \times \mathcal{Y}^n)$ is bounded by $(n+1)^{|\mathcal{X}||\mathcal{Y}|}$ and (82) is satisfied. Therefore, for a given source sequence x^n and side information sequence y^n , the henchman is able to send a message at rate $(R_L - R_0) + \varepsilon$ such that the eavesdropper can produce z^n with distortion

$$d_E(x^n, z^n) \leq D(R_L - R_0 + \varepsilon, T_{x^n y^n}) + \varepsilon. \quad (83)$$

Now define

$$Q_{XY}^* \triangleq \arg \max_{Q \in \mathcal{A}_n} D(R_L - R_0 + \varepsilon, Q). \quad (84)$$

From (81), we see that with probability at least ε , the henchman and the eavesdropper can achieve distortion

$$d_E(X^n, Z^n) \leq D(R_L - R_0 + \varepsilon, T_{X^n Y^n}) + \varepsilon \quad (85)$$

$$\leq D(R_L - R_0 + \varepsilon, Q_{XY}^*) + \varepsilon \quad (86)$$

Therefore, in view of (80), we can bound D_E :

$$D_E \stackrel{(a)}{\leq} D(R_L - R_0 + \varepsilon, Q_{XY}^*) + 2\varepsilon \quad (87)$$

$$\stackrel{(b)}{\leq} D(R_L - R_0 + \varepsilon, P_X Q_{Y|X}^*) + 2\varepsilon + o(\varepsilon). \quad (88)$$

Step (a) follows from (80). Step (b) is due to $\|Q_X^* - P_X\|_{TV} < \varepsilon$ and the fact that the rate-distortion function is continuous in P_X with respect to total variation distance (e.g., see [12]). Because $Q_{XY}^* \in \mathcal{A}_n$, we can also bound R and D_B . First, we have

$$R \geq I(Q_{XY}^*) - \varepsilon \quad (89)$$

$$\stackrel{(a)}{\geq} I(P_X Q_{Y|X}^*) - \varepsilon - o(\varepsilon), \quad (90)$$

where (a) is due to the continuity of mutual information with respect to total variation distance. Next, we have

$$D_B \geq \mathbb{E}_{Q_{XY}^*} d_B(X, Y) - \varepsilon \quad (91)$$

$$\stackrel{(a)}{\geq} \mathbb{E}_{P_X Q_{Y|X}^*} d_B(X, Y) - \varepsilon - o(\varepsilon), \quad (92)$$

where (a) uses Property 1c of total variation.

Assimilating the bounds that we have established, we can conclude that any achievable tuple (R, R_0, R_L, D_B, D_E) lies in the region

$$\mathcal{S}_\varepsilon \triangleq \bigcup_{P_{Y|X}} \left\{ (R, R_0, R_L, D_B, D_E) : \begin{aligned} R &\geq I(X; Y) - o(\varepsilon) \\ D_B &\geq \mathbb{E} d_B(X, Y) - o(\varepsilon) \\ D_E &\leq \min\{D(R_L), D(R_L - R_0 + \varepsilon, P_{Y|X}) + o(\varepsilon)\} \end{aligned} \right\}. \quad (93)$$

Since this holds for all $\varepsilon > 0$, we have

$$(R, R_0, R_L, D_B, D_E) \in \bigcap_{\varepsilon > 0} \mathcal{S}_\varepsilon. \quad (94)$$

The region in (94) is equal to the region in Theorem 3 (subject to $R_L \geq R_0$), completing the converse proof.

VIII. ACHIEVABILITY (LOSSY COMMUNICATION)

In this section, we prove the achievability portion of Theorem 3, the lossy communication counterpart to Theorem 1. The skeleton of the proof is similar to the one presented in Section V, but we will need some enhanced versions of some of the components.

As in the lossless setting, we can view the henchman and the eavesdropper as the sender and receiver in a rate-limited system with side information M (i.e., the public message) available to both parties. The correlation between the side information and the source sequence X^n will govern the performance; therefore, we are interested in $P_{X^n|M=m}$ since this is the effective source distribution after accounting for common side information. As before, the encoder at Node A determines $P_{X^n|M=m}$ entirely. In Section V, we were motivated by the effect of random binning (which we later replaced with a likelihood encoder for ease of analysis). However, instead of simply randomly binning X^n and using K to hide the location within the bin, we now want to first perform lossy compression using a codebook of sequences from \mathcal{Y}^n , followed by a random binning of the codebook. Roughly speaking, this process results in a distribution $P_{X^n|M=m}$ that corresponds to selecting a y^n sequence uniformly from a random codebook of size 2^{nR_0} , then passing that sequence through a memoryless channel $\prod P_{X|Y}$. The justification for this assertion will become clear when we use a likelihood encoder later on; for now, we study the subproblem that just surfaced: lossy compression of a *noisy version* of a codeword drawn uniformly from a random codebook.

A. Lossy compression of a noisy version of a codeword drawn uniformly from a random codebook

Consider a codebook $c_y = \{y^n(1), \dots, y^n(2^{nR_0})\}$ consisting of 2^{nR_0} sequences in \mathcal{Y}^n . Select a codeword uniformly at random from c_y and denote it by $y^n(J)$, where $J \sim \text{Unif}[2^{nR_0}]$. Pass $y^n(J)$ through a memoryless channel $\prod P_{X|Y}$ to produce a sequence X^n . An encoder describes X^n using a noiseless link of rate R , and a decoder

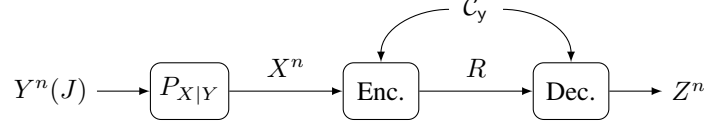


Fig. 6: Lossy compression of a noisy version of a codeword drawn uniformly from a random codebook $\mathcal{C}_y = \{Y^n(1), \dots, Y^n(2^{nR_c})\}$. Both the encoder and decoder know the codebook \mathcal{C}_y . The encoder describes X^n , the output of a memoryless channel $\prod P_{X|Y}$ whose input is a randomly chosen codeword $Y^n(J)$, where $J \sim \text{Unif}[2^{nR_c}]$.

estimates it with a reconstruction sequence Z^n (incurring distortion $d(X^n, Z^n)$). Both the encoder and decoder know the codebook c_y , and together they constitute a (n, c_y, R) code. The setup is shown in Figure 6 for a random codebook \mathcal{C}_y .

The following theorem generalizes Theorem 2 (to recover that theorem, set $Y = X$).

Theorem 4. Fix P_{XY} , R , R_c and D . Let \mathcal{C}_y be a random codebook of 2^{nR_c} codewords, each drawn independently according to $\prod_{i=1}^n P_Y(y_i)$. Let τ_n be any sequence that converges to zero sub-exponentially fast (i.e., $\tau_n = 2^{-o(n)}$). If

$$R < \min\{R(D), R_Y(D) + R_c\}, \quad (95)$$

then with high probability it is impossible to achieve distortion D in the sense that

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\mathcal{C}_y} \left[\max_{(n, \mathcal{C}_y, R) \text{ codes}} \mathbb{P}[d(X^n, Z^n) \leq D] > \tau_n \right] = 0. \quad (96)$$

The function $R_Y(D)$ is the rate-distortion function with side information:

$$R_Y(D) = \min_{P_{Z|XY} : \mathbb{E} d(X, Z) \leq D} I(X; Z|Y). \quad (97)$$

Before diving into the proof, let us briefly justify why the regime in (95) is the one of interest. First, observe that whenever $R \geq R(D)$ is satisfied, distortion D can be achieved by simply using a regular point-to-point rate distortion code to describe X^n . Second, whenever $R \geq R_Y(D) + R_c$ holds, distortion D can be achieved in roughly the following manner. The encoder first identifies a codeword in \mathcal{C}_y that is jointly typical with X^n (according to P_{XY}) and sends the index of the codeword using rate R_c . The codeword is then treated as side information, which allows the encoder to describe X^n using rate $R_Y(D)$. So we see that (95) is actually necessary for (96) to hold.

Proof: We follow the basic rubric of Section V, making modifications where they are needed.

Fixing P_{XY} , we first restrict $(X^n, Y^n(J))$ to be jointly typical by writing

$$\mathbb{P}[d(X^n, Z^n) \leq D] \leq \mathbb{P}[d(X^n, Z^n) \leq D, \mathcal{A}] + \mathbb{P}[\mathcal{A}^c], \quad (98)$$

where \mathcal{A} denotes the event $\{(X^n, Y^n(J)) \in \mathcal{T}_\delta^n(X, Y)\}$. Note that the second term in (98) vanishes in the limit for any $\delta > 0$ since $(X^n, Y^n(J))$ is i.i.d. according to P_{XY} .

Continuing exactly as in Section V, we have

$$\max_{(n, C_Y, R) \text{ codes}} \mathbb{P}[d(X^n, Z^n) \leq D, \mathcal{A}] \leq 2^{nR} \max_{z^n \in \mathcal{Z}^n} \mathbb{P}[d(X^n, z^n) \leq D, \mathcal{A}] \quad (99)$$

$$= 2^{nR} \max_{z^n \in \mathcal{Z}^n} \mathbb{E}_J \mathbb{P}[d(X^n, z^n) \leq D, \mathcal{A} | Y^n(J)] \quad (100)$$

$$= 2^{-n(R_C - R)} \max_{z^n \in \mathcal{Z}^n} \sum_{j=1}^{2^{nR_C}} \mathbb{P}[d(X^n, z^n) \leq D, \mathcal{A} | Y^n(j)] \quad (101)$$

Denote the terms in the sum by ζ_{j, z^n} :

$$\zeta_{j, z^n} \triangleq \mathbb{P}[d(X^n, z^n) \leq D, \mathcal{A} | Y^n(j)] \quad (102)$$

$$= \sum_{x^n \in \mathcal{X}^n} \prod_{i=1}^n P_{X|Y}(x_i | Y_i(j)) \cdot \mathbf{1}\{d(x^n, z^n) \leq D, (x^n, Y^n(j)) \in \mathcal{T}_\delta\} \quad (103)$$

Continuing in the manner of Section V leads us to

$$\mathbb{P}\left[\max_{(n, C_Y, R) \text{ codes}} \mathbb{P}[d(X^n, Z^n) \leq D, \mathcal{A}] > \tau_n\right] \leq |\mathcal{Z}|^n \max_{z^n \in \mathcal{Z}^n} \mathbb{P}\left[\sum_{j=1}^{2^{nR_C}} \zeta_{j, z^n} > \tau_n 2^{n(R_C - R)}\right], \quad (104)$$

As with the ξ_{j, z^n} defined in Section V (Eq. (35)), the ζ_{j, z^n} are i.i.d. due to the nature of the random codebook; however, they are no longer Bernoulli random variables. The following lemma, a straightforward generalization of Lemma 1, shows that ζ_{j, z^n} is bounded above by $2^{-n(R_Y(D) - o(1))}$ with probability one. The proof is omitted.

Lemma 6. Fix P_{XY} and $y^n \in \mathcal{Y}^n$. If X^n is distributed according to $\prod_{i=1}^n P_{X|Y=y_i}$, then for any z^n ,

$$\mathbb{P}[d(X^n, z^n) \leq D, (X^n, y^n) \in \mathcal{T}_\delta^n | Y^n = y^n] \leq 2^{-n(R_Y(D) - o(1))}, \quad (105)$$

where $o(1)$ is a term that vanishes as $\delta \rightarrow 0$ and $n \rightarrow \infty$.

As mentioned, Lemma 6 implies

$$\zeta_{j, z^n} \in [0, 2^{-n(R_Y(D) - o(1))}]. \quad (106)$$

In addition to bounding the range of ζ_{j, z^n} we can also bound its expected value. In fact, the bound is the same as for ξ_{j, z^n} .

$$\mathbb{E}_{C_Y} \zeta_{j, z^n} = \mathbb{E}_{C_Y} \mathbb{P}[d(X^n, z^n) \leq D, \mathcal{A} | Y^n(j)] \quad (107)$$

$$\leq \mathbb{E}_{C_Y} \mathbb{P}[d(X^n, z^n) \leq D, X^n \in \mathcal{T}_\delta | Y^n(j)] \quad (108)$$

$$= \mathbb{P}[d(X^n, z^n) \leq D, X^n \in \mathcal{T}_\delta] \quad (\text{where } X^n \sim \prod P_X) \quad (109)$$

$$\stackrel{(a)}{\leq} 2^{-n(R(D) - o(1))}, \quad (110)$$

where (a) is due to Lemma 1.

We are now ready to apply a Chernoff bound to the probability in (104). First, we extend the Chernoff bound in Lemma 2 to random variables taking values on the interval $[0, a]$ instead of just binary random variables.

Corollary 2. *If X^m is a sequence of i.i.d. random variables on the interval $[0, a]$ with $\mathbb{E}[X_i] = p$, then*

$$\mathbb{P}\left[\sum_{i=1}^m X_i > k\right] \leq \left(\frac{e \cdot m \cdot p}{k}\right)^{k/a}. \quad (111)$$

Proof: We start by proving the case $a = 1$. To begin, we claim that if $X \in [0, 1]$ and $Y \in \{0, 1\}$ are random variables such that $\mathbb{E}[X] = \mathbb{E}[Y]$ and $f : [0, 1] \rightarrow \mathbb{R}$ is convex, then

$$\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]. \quad (112)$$

To see this, observe that for $x \in [0, 1]$,

$$f(x) \leq xf(1) + (1 - x)f(0). \quad (113)$$

Taking expectations gives

$$\mathbb{E}[f(X)] \leq \mathbb{E}[X]f(1) + (1 - \mathbb{E}[X])f(0) \quad (114)$$

$$= \mathbb{E}[Y]f(1) + (1 - \mathbb{E}[Y])f(0) \quad (115)$$

$$= \mathbb{E}[f(Y)], \quad (116)$$

verifying the claim. Now, since $f(x) = e^{\lambda x}$ is convex, the inequality $\mathbb{E}[e^{\lambda Y}] \leq \mathbb{E}[e^{\lambda X}]$ holds and can be applied to the proof of Lemma 2 at (40).

With the case $a = 1$ shown, we now consider any $a > 0$. If we let $Y_i = \frac{1}{a}X_i \in [0, 1]$, then the previous case applies and we have

$$\mathbb{P}\left[\sum_{i=1}^m X_i > k\right] \leq \mathbb{P}\left[\sum_{i=1}^m aY_i > \frac{k}{a}\right] \quad (117)$$

$$\leq \left(\frac{e \cdot m \cdot \mathbb{E}[Y_1]}{k/a}\right)^{k/a} \quad (118)$$

$$= \left(\frac{e \cdot m \cdot p}{k}\right)^{k/a}. \quad (119)$$

■

Using the support bound and the expected value bound in (106) and (110), we can apply Corollary 2 to the probability in (104) by identifying

$$m = 2^{nR_c} \quad (120)$$

$$a = 2^{-n(R_Y(D) - o(1))} \quad (121)$$

$$p \leq 2^{-n(R(D) - o(1))} \quad (122)$$

$$k = \tau_n 2^{n(R_c - R)}. \quad (123)$$

This gives

$$\mathbb{P}\left[\sum_{j=1}^{2^{nR_C}} \zeta_{j,z^n} > \tau_n 2^{n(R_C-R)}\right] \leq 2^{-n\alpha 2^{n\beta}}, \quad (124)$$

where

$$\alpha = R(D) - R - o(1) \quad (125)$$

$$\beta = R_C + R_Y(D) - R - o(1). \quad (126)$$

For small enough δ and large enough n , both α and β are positive and bounded away from zero, and (124) vanishes doubly exponentially fast. Consequently, the expression in (104) vanishes, completing the proof of Theorem 4. ■

The following corollary to Theorem 4 is immediate, and, as in Section V, will serve as the bridge between the subproblem we have been considering and the henchman problem.

Corollary 3. Fix P_{XY} . If $R < \min\{R(D), R_Y(D) + R_C\}$, then

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}_Y} \left[\min_{(n, \mathcal{C}_Y, R) \text{ codes}} \mathbb{P}[d(X^n, Z^n) \geq D] \right] = 1. \quad (127)$$

B. Likelihood encoder

Returning to the henchman problem, we follow the basic structure of Section V. Fixing $P_{Y|X}$ (and thus a joint distribution P_{XY}), consider a codebook $c = \{y^n(m, k)\}$ of $2^{n(R+R_0)}$ sequences from \mathcal{Y}^n and define a likelihood encoder for this codebook by

$$P_{M|X^n K}(m|x^n, k) \propto \prod_{i=1}^n P_{X|Y}(x_i|y_i(m, k)), \quad (128)$$

where \propto indicates that appropriate normalization is required. The distribution $P_{X^n M K}$ induced by using this encoder with a random codebook is intimately related to an idealized distribution $Q_{X^n M K}$ defined by

$$Q_{X^n M K}(x^n, m, k) \triangleq 2^{-n(R+R_0)} \prod_{i=1}^n P_{X|Y}(x_i|y_i(m, k)). \quad (129)$$

Indeed, just as in Lemma 3, one can use the soft covering lemma to show that if $R > I(X; Y)$, then

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}} \|P_{X^n M K} - Q_{X^n M K}\|_{\text{TV}} = 0, \quad (130)$$

where \mathcal{C} is a random codebook with each codeword drawn independently according to $\prod_{i=1}^n P_Y$.

Inspecting $Q_{X^n M K}$ reveals that $Q_{X^n|M=m}$ is exactly the distribution that was addressed in the recent interlude. To see this, observe that $Q_{X^n K|M=m}$ is the joint distribution that arises from selecting a codeword uniformly from a codebook of size 2^{nR_0} and passing it through a memoryless channel $\prod P_{X|Y}$. To be explicit,

$$Q_{X^n|M}(x^n|m) = 2^{-nR_0} \sum_{k=1}^{2^{nR_0}} \prod_{i=1}^n P_{X|Y}(x_i|Y_i(m, k)). \quad (131)$$

Proceeding with the analysis of the eavesdropper's distortion, first denote the event

$$\mathcal{E} = \{d(X^n, z^n(M, M_H)) \geq \pi(R_L, R_0, P_{Y|X}) - \varepsilon\}, \quad (132)$$

where

$$\pi(R_L, R_0, P_{Y|X}) \triangleq \begin{cases} D(R_L) & \text{if } R_0 > R_L \\ \min\{D(R_L), D_Y(R_L - R_0)\} & \text{if } R_0 \leq R_L \end{cases} \quad (133)$$

The purpose of $\pi(\cdot)$ is to treat the cases $R_L < R_0$ and $R_L \geq R_0$ concurrently.

Taking the expectation of (68) with respect to a random codebook, we have

$$\mathbb{E}_{\mathcal{C}} \left[\min_{\substack{m_H(x^n, m), z^n(m, m_H): \\ |\mathcal{M}_H| \leq 2^{nR_L}}} \mathbb{P}_{P_{X^n M}}[\mathcal{E}] \right] = \mathbb{E}_M \mathbb{E}_{\mathcal{C}} \left[\min_{\substack{m_H(x^n), z^n(m_H): \\ |\mathcal{M}_H| \leq 2^{nR_L}}} \mathbb{P}_{Q_{X^n|M}}[\mathcal{E}|M] \right] + o(1). \quad (134)$$

From (131), we see that the expression in (134) is exactly what is addressed by Corollary 3 after identifying (R_0, R_L) with (R_C, R) . Note that we are invoking the corollary with $D = \pi(R_L, R_0, P_{Y|X}) - \varepsilon$, which means that

$$R_L < \min\{R(D), R_Y(D) + R_0\}. \quad (135)$$

Thus, we have

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}} \left[\min_{\substack{m_H(x^n), z^n(m_H): \\ |\mathcal{M}_H| \leq 2^{nR_L}}} \mathbb{P}_{Q_{X^n|M=m}}[\mathcal{E}|M=m] \right] = 1. \quad (136)$$

Therefore, we can conclude that there exists a codebook such that the associated likelihood encoder ensures (68), because (68) holds when averaged over random codebooks.

We now complete the proof of achievability by showing that the likelihood encoder can be used to achieve distortion $\mathbb{E} d_B(X, Y)$ at the legitimate receiver (this is also done in [7]). To do this, Node B uses a deterministic decoder that simply produces the codeword indexed by (m, k) , i.e.,

$$P_{Y^n|MK}(y^n|m, k) = \mathbf{1}\{y^n = y^n(m, k)\}. \quad (137)$$

Defining $Q_{X^n MKY^n} \triangleq Q_{X^n MK} P_{Y^n|MK}$, we can write

$$\mathbb{E}_{\mathcal{C}} \|P_{X^n Y^n} - Q_{X^n Y^n}\|_{\text{TV}} \stackrel{(a)}{\leq} \mathbb{E}_{\mathcal{C}} \|P_{X^n MK} P_{Y^n|MK} - Q_{X^n MK} P_{Y^n|MK}\|_{\text{TV}} \quad (138)$$

$$\stackrel{(b)}{=} \mathbb{E}_{\mathcal{C}} \|P_{X^n MK} - Q_{X^n MK}\|_{\text{TV}} \quad (139)$$

$$\stackrel{(c)}{\rightarrow} 0, \quad (140)$$

where (a) and (b) follow from Properties 1d and 1c, and (c) is due to (130). Now notice that $\mathbb{E}_{\mathcal{C}} Q_{X^n Y^n}$ is exactly the product distribution $\prod_{i=1}^n P_{XY}$ (a fact which is straightforward to verify). Therefore, by (140) and the weak law of large numbers, we have

$$\mathbb{E}_{\mathcal{C}} \mathbb{P} \left[d_B(X^n, Y^n) > \mathbb{E} d_B(X, Y) + \varepsilon \right] = \mathbb{E}_{\mathcal{C}} \mathbb{P}_{Q_{X^n Y^n}} \left[d_B(X^n, Y^n) > \mathbb{E} d_B(X, Y) + \varepsilon \right] + o(1) \quad (141)$$

$$= o(1) \quad (142)$$

This completes the achievability portion of the proof of Theorem 3.

APPENDIX A
PROOF OF LEMMA 1

We first bound $\mathbb{P}[d(X^n, z^n) \leq D]$ and resolve the event $X^n \in \mathcal{T}_\delta^n$ afterward. We use the V-shell notation from the method of types [9]: for a stochastic matrix $V_{Z|X}$, the set of z^n sequences having conditional type V is denoted by $\mathcal{T}_V^n(x^n)$. Note that all pairs (x^n, z^n) satisfying $z^n \in \mathcal{T}_V^n(x^n)$ have the same joint type (denoted by $T_{x^n z^n}$).

Diving in, we have

$$\mathbb{P}[d(X^n, z^n) \leq D] = \sum_{x^n \in \mathcal{X}^n} P_{X^n}(x^n) \mathbf{1}\{d(x^n, z^n) \leq D\} \quad (143)$$

$$\stackrel{(a)}{=} \sum_{V_{X|Z}} \sum_{x^n \in \mathcal{T}_V^n(z^n)} P_{X^n}(x^n) \mathbf{1}\{d(x^n, z^n) \leq D\} \quad (144)$$

$$\stackrel{(b)}{=} \sum_{V_{X|Z}} \sum_{x^n \in \mathcal{T}_V^n(z^n)} 2^{-n(D(T_{x^n} \| P_X) + H(T_{x^n}))} \mathbf{1}\{\mathbb{E}_{T_{x^n z^n}} d(X, Z) \leq D\} \quad (145)$$

$$\stackrel{(c)}{\leq} \sum_{\substack{V_{X|Z}: \\ \mathcal{T}_V^n(z^n) \neq \emptyset}} 2^{nH(T_{x^n} | T_{z^n})} 2^{-n(D(T_{x^n} \| P_X) + H(T_{x^n}))} \mathbf{1}\{\mathbb{E}_{T_{x^n z^n}} d(X, Z) \leq D\} \quad (146)$$

$$= \sum_{\substack{V_{X|Z}: \\ \mathcal{T}_V^n(z^n) \neq \emptyset}} 2^{-n(I(T_{x^n z^n}) + D(T_{x^n} \| P_X))} \mathbf{1}\{\mathbb{E}_{T_{x^n z^n}} d(X, Z) \leq D\} \quad (147)$$

$$\stackrel{(d)}{\leq} \exp \left\{ -n \min_{V: \mathbb{E}_{T_{x^n z^n}} d(X, Z) \leq D} [I(T_{x^n z^n}) + D(T_{x^n} \| P_X)] + O(\log n) \right\}. \quad (148)$$

In step (a), we partition the set \mathcal{X}^n according to the conditional type of x^n given z^n . Step (b) follows by observing that the summands only depend on the joint type of (x^n, z^n) . Step (c) uses a bound on the size of $\mathcal{T}_V^n(z^n)$, and step (d) follows from the fact that the number of conditional types is polynomial in n .

We can continue by lower bounding the first term in the (normalized) exponent of (148):

$$\begin{aligned} & \min_{V: \mathbb{E}_{T_{x^n z^n}} d(X, Z) \leq D} I(T_{x^n z^n}) + D(T_{x^n} \| P_X) \\ & \geq \min_{z^n} \min_{V: \mathbb{E}_{T_{x^n z^n}} d(X, Z) \leq D} I(T_{x^n z^n}) + D(T_{x^n} \| P_X) \end{aligned} \quad (149)$$

$$= \min_{Q_{XZ}: \mathbb{E}_Q d(X, Z) \leq D} I_Q(X; Z) + D(Q_X \| P_X) \quad (150)$$

$$= \min_{Q_X} \min_{Q_{Z|X}: \mathbb{E}_Q d(X, Z) \leq D} I_Q(X; Z) + D(Q_X \| P_X) \quad (151)$$

$$= \min_{Q_X} [R(D, Q_X) + D(Q_X \| P_X)], \quad (152)$$

where $R(D, Q_X)$ denotes the rate-distortion function for a source Q_X .

So far, we have shown that the following holds for all z^n :

$$\mathbb{P}[d(X^n, z^n) \leq D] \leq \exp \{ -n \cdot \min_{Q_X} [R(D, Q_X) + D(Q_X \| P_X)] + O(\log n) \}. \quad (153)$$

However, this is not quite the bound we seek; a simple example will reveal that it is possible to have

$$\min_{Q_X} [R(D, Q_X) + D(Q_X || P_X)] < R(D). \quad (154)$$

Indeed, consider $P_X \sim \text{Bern}(p)$, $p \in (D, 1/2)$ and $Q_X \sim \text{Bern}(q)$. After simplifying, we find that

$$R(D, Q_X) + D(Q_X || P_X) = q \log \frac{1}{p} + (1-q) \log \frac{1}{1-p} - h(D). \quad (155)$$

Minimizing this expression over $q \in [0, 1]$ gives

$$\min_{Q_X} [R(D, Q_X) + D(Q_X || P_X)] = \min \left\{ \log \frac{1}{p}, \log \frac{1}{1-p} \right\} - h(D) \quad (156)$$

$$< h(p) - h(D) \quad (157)$$

$$= R(D). \quad (158)$$

To resolve this issue, we introduce the event $X^n \in \mathcal{T}_\delta^n$ into the expression we want to bound. Modifying the steps above accordingly, we have

$$-\frac{1}{n} \log \mathbb{P}[d(X^n, z^n) \leq D, X^n \in \mathcal{T}_\delta^n] \geq \min_{Q_X: \|Q_X - P_X\|_{\text{TV}} < \delta} R(D, Q_X) + D(Q_X || P_X) - O\left(\frac{\log n}{n}\right) \quad (159)$$

$$\stackrel{(a)}{\geq} \min_{Q_X: \|Q_X - P_X\|_{\text{TV}} < \delta} R(D, Q_X) - O\left(\frac{\log n}{n}\right) \quad (160)$$

$$\stackrel{(b)}{=} R(D) - O\left(\delta \log \frac{1}{\delta}\right) - O\left(\frac{\log n}{n}\right) \quad (161)$$

$$= R(D) - o(1), \quad (162)$$

where step (a) is due to the non-negativity of relative entropy and step (b) follows from the uniform continuity of the rate-distortion function with respect to total variation distance (e.g., [12]).

REFERENCES

- [1] C. Schieler and P. Cuff, “The henchman problem: measuring secrecy by the minimum distortion in a list,” in *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, Jul. 2014.
- [2] C. Shannon, “Communication theory of secrecy systems,” *Bell Syst. Tech. J.*, vol. 28, no. 4, pp. 656–715, Oct. 1949.
- [3] H. Yamamoto, “Rate-distortion theory for the Shannon cipher system,” *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 827–835, May 1997.
- [4] C. Schieler and P. Cuff, “Rate-distortion theory for secrecy systems,” *To appear in IEEE Trans. Inf. Theory*, 2014. [Online]. Available: <http://arxiv.org/abs/1305.3905>
- [5] N. Merhav and E. Arikan, “The Shannon cipher system with a guessing wiretapper,” *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1860–1866, 1999.
- [6] E. Haroutunian, “On the Shannon cipher system with a wiretapper guessing subject to distortion and reliability requirements,” Aug. 2010. [Online]. Available: <http://arxiv.org/abs/1008.0961>
- [7] E. C. Song, P. Cuff, and H. V. Poor, “The likelihood encoder for lossy source compression,” Apr. 2014. [Online]. Available: <http://arxiv.org/abs/1404.5683>
- [8] P. Cuff, “Distributed channel synthesis,” *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, 2013.
- [9] I. Csiszár, “The method of types,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, 1998.
- [10] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

- [11] T. Weissman and E. Ordentlich, “The empirical distribution of rate-constrained source codes,” *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3718–3733, Nov. 2005.
- [12] H. Palaiyanur and A. Sahai, “On the uniform continuity of the rate-distortion function,” in *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, Jul. 2008, pp. 857–861.